

caIntegrator

End User Research Portals



U.S. DEPARTMENT
OF HEALTH AND
HUMAN SERVICES

National Institutes
of Health



Agenda

- calIntegrator Background
- Current Status
- Future Directions
- Questions

What is caIntegrator

- Brief Description

- caIntegrator is a web-based software application that allows researchers to set up custom, caBIG®-compatible web portals to conduct integrative research, without requiring programming experience.
- These portals bring together heterogeneous clinical, microarray and medical imaging data to enrich multidisciplinary research.
- NCI-hosted caIntegrator provides access to public data including,
 - REMBRANDT (Repository for Molecular Brain Neoplasia Data)
 - TARGET-ALL (Therapeutically Applicable Research to Generate Effective Treatments- Acute Lymphoblastic Leukemia)
 - TCGA (The Cancer Genome Atlas)

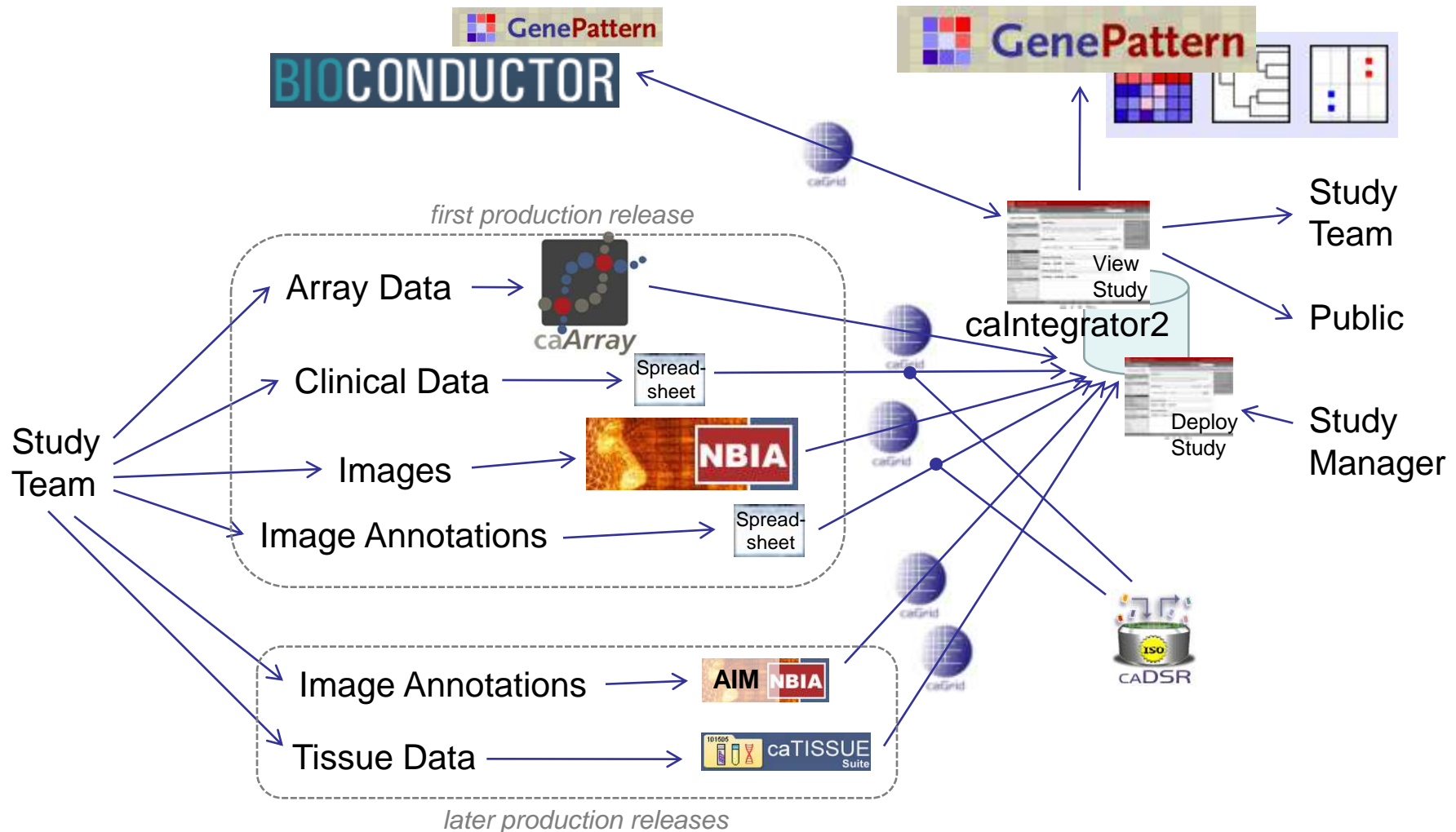
- Intended Audiences:

- Basic and translational researcher

calIntegrator Goals

- Have a single place where one can go to get all the information about a study
 - Clinical Data
 - Imaging Data
 - Genomic Data
- Provide ways to browse, query, visualize and analyze the data
 - Explore and confirm hypotheses
 - Share data and results, e.g. lists of interesting genes
- Have a way to roll out new studies easily
- Have a consistent user experience across different studies
- Lay the groundwork for cross-study comparisons

How Does caIntegrator2 Work?



calIntegrator

Aggregate Data, Create a Study

PATIENT_ID	MICROARRAY	SITE	GENDER	AGE AT DIAGNOSIS	RACE	SMOKING_HISTORY	PATHOLOGIC_N_STAGE
151	NCI_U133A_1L	MSKCC	Male	64	White(01)	Currently smoking	N1, ACCORDING TO AJCC CRITERIA
152	NCI_U133A_2L	MSKCC	Male	62	White(01)	Smoked in the past	N0, ACCORDING TO AJCC CRITERIA
153	NCI_U133A_3L	MSKCC	Female	70	White(01)	Smoked in the past	N0, ACCORDING TO AJCC CRITERIA
154	NCI_U133A_4L	MSKCC	Male	67	White(01)	Smoked in the past	N2, ACCORDING TO AJCC CRITERIA
155	NCI_U133A_5L	MSKCC	Female	56	White(01)	Currently smoking	N1, ACCORDING TO AJCC CRITERIA
156	NCI_U133A_6L	MSKCC	Male	57	White(01)	Smoked in the past	N1, ACCORDING TO AJCC CRITERIA
157	NCI_U133A_7L	MSKCC	Female	81	White(01)	Never smoked	N0, ACCORDING TO AJCC CRITERIA
158	NCI_U133A_8L	MSKCC	Female	65	White(01)	Smoked in the past	N0, ACCORDING TO AJCC CRITERIA
159	NCI_U133A_9L	MSKCC	Female	73	White(01)	Never smoked	N2, ACCORDING TO AJCC CRITERIA
160	NCI_U133A_10L	MSKCC	Male	65	White(01)	Currently smoking	N0, ACCORDING TO AJCC CRITERIA
161	NCI_U133A_11L	MSKCC	Male	62	White(01)	Smoked in the past	N0, ACCORDING TO AJCC CRITERIA
162	NCI_U133A_12L	MSKCC	Male	71	White(01)	Smoked in the past	N2, ACCORDING TO AJCC CRITERIA
163	NCI_U133A_13L	MSKCC	Female	82	White(01)	Currently smoking	N0, ACCORDING TO AJCC CRITERIA
164	NCI_U133A_14L	MSKCC	Male	68	White(01)	Smoked in the past	N0, ACCORDING TO AJCC CRITERIA

Clinical data file saved from excel in CSV format

Creating and Managing Studies

The screenshot displays the 'Edit Study' interface in the caIntegrator2 application. The page is divided into a left sidebar with navigation links and a main content area. The main content area includes sections for 'Clinical Data Sources', 'Genomics Data Sources', and 'Imaging Data Sources'. Each section contains a table of configured data sources with columns for Host Name, Experiment Identifier, File Description, Status, and Action.

Study Details

Study Name: TCGA Agilent
Study Description: TCGA GBM Agilent with image mappings
Status: Deployed
Status Description: Minutes for deployment (approx): 265
Buttons: Save, Save and Deploy, Change Study Logo

Clinical Data Sources

Type	Description
DELIMITED_TEXT	logs_gbm_public_clinical.csv

File type (CSV only):

Genomics Data Sources

Host Name	Experiment Identifier	File Description	Status	Action
array.nci.nih.gov	tcga-00047	Mapping File(s): array_mappings_agilent_tcga_47.csv Control Sample Mapping File(s): None Configured Copy Number Mapping File: None Configured		
array.nci.nih.gov	tcga-00050	Mapping File(s): array_mappings_agilent_tcga_50.csv Control Sample Mapping File(s): None Configured Copy Number Mapping File: None Configured		

Imaging Data Sources

Host Name	Collection Name	File Description
imaging.nci.nih.gov	TCGA	Annotation: None Mapping: tcga_gbm_image_mappings.csv

Querying Study Data



National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov

caINTEGRATOR2

build: CAINTEGRATOR2_R1_0_3_GA | date: '10/15/2009
02:14 PM

My Studies: **DC Lung Study demo K**

Welcome, [ketchumka](#) | [Logout](#)

Sample Study Logo

(max. 200px x 72px)

DC LUNG STUDY DEMO KAK

[Home](#)

[Search DC Lung Study demo KAK](#)

STUDY DATA

[Queries](#)

ANALYSIS TOOLS

[KM Plot](#)

[Gene Expression Plot](#)

[GenePattern Analysis](#)

STUDY MANAGEMENT

[Manage Studies](#)

[Create New Study](#)

APPLICATION MANAGEMENT

[Manage Platforms](#)

caINTEGRATOR2 MENU

[Support](#)

[Tutorials](#)

[User Guide](#)

Search DC Lung Study demo KAK

Criteria

Results Type

Sorting

Query Results

Save as...

Define Query Criteria for: Unsaved Query

Clinical

[+ Add](#)

Clinical

☐ or ☒ and

- Please Select --
- Please Select --
- *Unique Identifier
- ADJUVANT_CHEMO
- ADJUVANT_RT
- AGE_AT_DIAGNOSIS
- DC_STUDY_ID
- DEATH_DATE
- ENROLLMENT_DATE
- FIRST_PROGRESSION_OR_RELAPSE
- GENDER
- Histologic grade
- IN_DC_STUDY
- LABORATORY_BATCH
- LAST_CLINICAL_ASSESSMENT_DATE
- LAST_CONTACT_DATE
- MEDIAN_INTENSITY_UNNORMALIZED
- MICROARRAY
- MONTHS_TO_FIRST_PROGRESSION
- MONTHS_TO_LAST_CONTACT_OR_DEATH
- MTHS_TO_LAST_CLINICAL_ASSESSMENT

[Remove](#)

[Run Query](#)

[CONTACT US](#)


[PRIVACY NOTICE](#)

[DISCLAIMER](#)

[ACCESSIBILITY](#)

[SUPPORT](#)

Querying Study Data

 National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov

caINTEGRATOR2
build: CAINTEGRATOR2_R1_0_3_GA | date: 10/15/2009
02:14 PM

My Studies: DC Lung Study demo K

Welcome, [ketchumka](#) | [Logout](#)

Sample Study Logo
(max. 200px x 72px)

DC LUNG STUDY DEMO KAK
[Home](#)
Search DC Lung Study demo KAK

STUDY DATA
▼ Queries
► My Queries

ANALYSIS TOOLS
[KM Plot](#)
[Gene Expression Plot](#)
[GenePattern Analysis](#)

STUDY MANAGEMENT
[Manage Studies](#)
[Create New Study](#)

APPLICATION MANAGEMENT
[Manage Platforms](#)

caINTEGRATOR2 MENU
[Support](#)
[Tutorials](#)
[User Guide](#)

Search DC Lung Study demo KAK

Criteria | Results Type | Sorting | Query Results | Save as...

Define Query Criteria for: Unsaved Query

Clinical + Add

Clinical	GENDER	equals	Female	Remove
Clinical	AGE_AT_DIAGNOSIS	in	<div>48 49 50 51 52</div>	Remove

☐ or ☒ and

Run Query

Querying Study Data

The screenshot shows the caIntegrator2 web interface. The browser address bar displays the URL: <https://caintegrator2-stage.nci.nih.gov/caintegrator2/manageQuery.action?selectedAction=createNewQuery>. The page header includes the National Cancer Institute logo and the text "U.S. National Institutes of Health | www.cancer.gov". The main content area is titled "Search DC Lung Study" and features tabs for "Criteria", "Columns", "Sorting", "Query Results", and "Save as...". A "Save This Query" dialog box is open, prompting the user to enter a "Query Name" and a "Query Description". A teal arrow points from the "Save as..." tab to the "Save This Query" dialog box. Another teal arrow points from the "Analysis Tools" menu to the "Gene Expression Plot" option. The left sidebar contains a navigation menu with sections: "The Director's Challenge", "DC LUNG STUDY", "STUDY DATA", "ANALYSIS TOOLS", "STUDY MANAGEMENT", "APPLICATION MANAGEMENT", and "caINTEGRATOR2 MENU".

Search DC Lung Study

Criteria Columns Sorting Query Results Save as...

Save This Query

Query Name:

Query Description:

Save

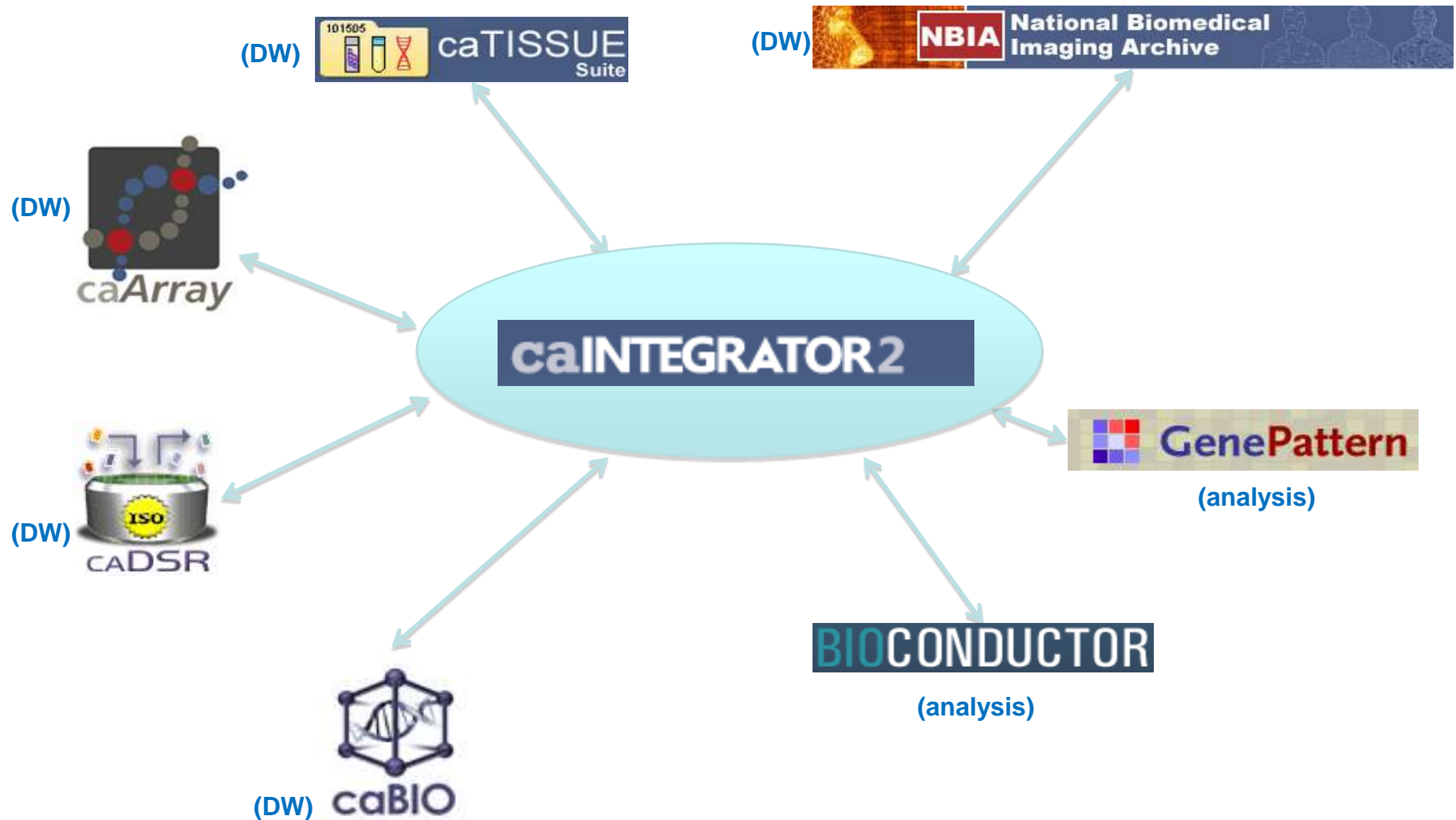
ANALYSIS TOOLS:

- KM Plot
- Gene Expression Plot
- GenePattern Analysis

Saved Queries are good for:

- Saving curated subsets of data
- Sharing those subsets with other users
- Storing subsets of data for use in other parts of caIntegrator2

Integration with Analysis Tools



calIntegrator2 1.1 Features

- List management: create, modify and delete lists, use of lists in queries and analysis.
- Pre-computed Genome Views: upload as cvs file or link out to pre-computed data.
- Enhancements to caBIO integration: partial match for gene search and pathway search.
- Support for Agilent gene expression data.
- Support for public browse option.
- Upgrade support: deploy new cal2 release without re-deploying the Studies.
- Enhancements to clinical data interface: large spreadsheets, easy upload of existing CDEs, easier annotation mapping.

Upcoming Features

- Biospecimen data from caTissue
- Time points
 - Clinical, array and imaging data can be associated with specific study time points (for instance, 'Time of Diagnosis' or 'Six Months after Treatment Start')
 - Queries will be able to operate on data from specific timepoints
- Support for Additional Data Types
 - SNP Queries
 - Methylation
 - Next Gen Sequencing support
- More analysis
 - Integration with other analysis tools such as geWorkbench and BioConductor

Q&A

Team Contact Information:

Mervi Heiskanen – heiskame@mail.nih.gov

Juli Klemm – klemmj@mail.nih.gov

Anand Basu – basuan@mail.nih.gov